

Mississippi Transportation Research Center

AUTOMATED ACCIDENT DETECTION AT INTERSECTIONS

Project No: FHWA/MS-DOT-RD-04-150



**U.S. Department
of Transportation
Federal Highway
Administration**



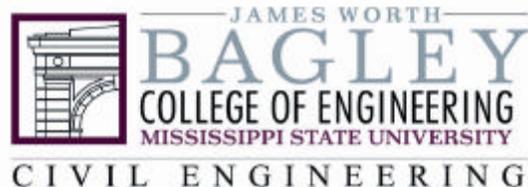
**Mississippi Department
of Transportation**

Prepared by:

**Yunlong Zhang
Department of Civil Engineering
Mississippi State University**

**Lori M. Bruce
Department of Electrical and Computer
Engineering
Mississippi State University**

March 2004



| | | | | | |
|---|--|--|--|---|--|
| 1. Report No. FHWA/MS-DOT-RD-04-150 | | 2. Government Accession No. | | 3. Recipient's Catalog No. | |
| 4. Title and Subtitle AUTOMATED ACCIDENT DETECTION AT INTERSECTIONS | | | | 5. Report Date March 2004 | |
| | | | | 6. Performing Organization Code | |
| 7. Author(s) Yunlong Zhang and Lori Mann Bruce | | | | 8. Performing Organization Report No. MS-DOT-RD-04-150 | |
| 9. Performing Organization Name and Address Department of Civil Engineering Mississippi State University P.O. Box 9546 Mississippi State, MS 39762-9546 | | | | 10. Work Unit No. (TRAIS) | |
| | | | | 11. Contract or Grant No. | |
| 12. Sponsoring Agency Name and Address Federal Highway Administration and Mississippi Department of Transportation | | | | 13. Type Report and Period Covered Final Report | |
| | | | | 14. Sponsoring Agency Code | |
| 15. Supplementary Notes Project conducted in cooperation with Federal Highway Administration and the Mississippi Department of Transportation | | | | | |
| 16. Abstract <p>This research aims to provide a timely and accurate accident detection method at intersections, which is very important for the Traffic Management System(TMS). This research uses acoustic signals to detect accident at intersections. A system is constructed that can be operated in two modes: two-class and multi-class. The input to the system is a three-second segment of audio signal. The output of the two-class mode is a label of "crash" or "non-crash". In the multi-class mode of operation, the system identifies crashes as well as several types of non-crash incidents, including normal traffic and construction sounds. The system is composed of three main signal processing stages: feature extraction, feature reduction, and feature classification. Five methods of feature extraction are investigated and compared; these are based on the discrete wavelet transform, fast Fourier transform, discrete cosine transform, real cepstral transform, and mel frequency cepstral transform. Statistical methods are used for feature optimization and classification. Three types of classifiers are investigated and compared: the nearest mean, maximum likelihood, and nearest neighbor methods. This study focuses on the detection algorithm development. Lab testing of the algorithm showed that the selected algorithm can detect intersection accidents with very high accuracy.</p> | | | | | |
| 17. Key Words Accident detection, Audio, Acoustic, Wavelet transform, Statistical classifier | | | | 18. Distribution Statement Unclassified | |
| 19. Security Classif. (of this report) Unclassified | | 20. Security Classif. (of this page) Unclassified | | 21. No. of Pages 66 | |
| | | | | 22. Price | |

ACKNOWLEDGMENT

The authors want to thank the Mississippi Department of Transportation (MDOT) for their sponsorship of the project. Special thanks are given to Mr. Bob Mabry of MDOT Traffic Engineering Division for his support and advice throughout the duration of this project. The authors also want to recognize the following graduate students for their contributions to the study and to this report:

| | |
|---------------------------|---|
| Navaneethakrishnan Balraj | Department of Electrical and Computer Engineering |
| Qingyong Yu | Department of Civil Engineering |
| Yuanchang Xie | Department of Civil Engineering |

TABLE OF CONTENTS

| | |
|---|-----------|
| ABSTRACT | vi |
| CHAPTER 1. INTRODUCTION | 1 |
| 1.1 Problem Statement | 1 |
| 1.2 Background | 3 |
| 1.3 Project Review | 5 |
| 1.4 Project Objective | 9 |
| CHAPTER 2. LITERATURE REVIEW | 10 |
| 2.1 Traffic Detectors | 10 |
| 2.1.1 Pressure Detector | 11 |
| 2.1.2 Inductive Loop Detector | 11 |
| 2.1.3 Passive Acoustic Detector | 12 |
| 2.1.4 Microwave Detector | 12 |
| 2.1.5 Passive Infrared Detector | 12 |
| 2.1.6 Video Detector | 13 |
| 2.1.7 Weather Detector | 13 |
| 2.2 Feature Extraction Technology | 13 |
| 2.2.1 Wavelet Analysis and DWT | 14 |
| 2.2.2 Fast Fourier Transform (FFT) | 16 |
| 2.2.3 Discrete Cosine Transform (DCT) | 17 |
| 2.2.4 Real Cepstral Transform (RCT) | 18 |
| 2.2.5 Mel Frequency Cepstral Transform (MCT) | 18 |
| 2.3 Feature Classification Methods | 20 |
| 2.3.1 Maximum Likelihood Classifier | 20 |
| 2.3.2 Nearest Mean Classifier | 20 |
| 2.3.3 Nearest Neighbor Classifier | 21 |
| CHAPTER 3. RESEARCH TASK AND PROCEDURE | 22 |
| 3.1 Data Collection and Processing | 22 |
| 3.1.1 Ordering Recording Equipment | 22 |
| 3.1.2 Data Collection | 23 |
| 3.1.3 Data Description and Preprocessing | 24 |
| 3.2 Feature Extraction | 31 |
| 3.2.1 Extraction Based on DWT | 32 |

| | |
|--|-----------|
| 3.2.2 Extraction Based on FFT | 33 |
| 3.2.3 Extraction Based on DCT | 34 |
| 3.2.4 Extraction Based on RCT | 34 |
| 3.2.5 Extraction Based on MCT..... | 35 |
| 3.3 Feature Reduction | 36 |
| 3.4 Feature Classification..... | 36 |
| 3.5 System Testing..... | 37 |
| CHAPTER 4.RESULTS AND DISCUSSION..... | 39 |
| CHAPTER 5.SUMMARY AND FUTURE ORIENTATIONS | 52 |
| 5.1 Summary of Testing Results and Recommendations | 52 |
| 5.2 Future Work | 53 |
| REFERENCES:..... | 56 |

LIST OF TABLES

| | |
|--|----|
| Table 4- 1 Maximum likelihood classification accuracies for two-class system..... | 41 |
| Table 4- 2 Maximum likelihood classification accuracies for multi-class system | 41 |
| Table 4- 3 Maximum likelihood classification accuracies..... | 43 |
| Table 4- 4 Overall accuracies of classification with DWT | 45 |
| Table 4- 5 Overall accuracies of classification with DWT | 45 |
| Table 4- 6 Maximum likelihood classification accuracies for various feature | 48 |
| Table 4- 7 Maximum likelihood classification accuracies for various feature | 48 |
| Table 4- 8 Classification Results with DWT and maximum likelihood classification | 51 |
| Table 4- 9 Classification Results with DWT and maximum likelihood classification | 51 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1- 1 Block diagram of automated accident detection system | 7 |
| Figure 3- 1 Crash sound in time domain and frequency domain | 27 |
| Figure 3- 2 Normal sound in time domain and frequency domain | 27 |
| Figure 3- 3 Synthesized crash sound (SNR=50dB) | 28 |
| Figure 3- 4 Synthesized crash sound (SNR=20dB) | 28 |
| Figure 3- 5 Synthesized crash sound (SNR=10dB) | 29 |
| Figure 3- 6 Synthesized crash sound (SNR=0dB) | 29 |
| Figure 3- 7 Synthesized crash sound (SNR=-10dB) | 30 |
| Figure 3- 8 Synthesized crash sound (SNR=-20dB) | 30 |
| Figure 3- 9 Synthesized crash sound (SNR=-50dB) | 31 |
| Figure 4- 1 Maximum likelihood classification accuracies | 43 |
| Figure 4- 2 DWT-based feature extraction using Haar mother wavelet | 46 |
| Figure 4- 3 DWT-based feature extraction using Haar mother wavelet | 46 |
| Figure 4- 4 Maximum likelihood classification accuracies for various feature | 49 |
| Figure 4- 5 Maximum likelihood classification accuracies for various feature | 49 |

ABSTRACT

Rapid increase of traffic demand has placed an increased strain on the already congested traffic system. It was estimated that the traffic congestion caused by incidents will cost the nation over \$75 billion in lost productivity by 2005. Among the types of traffic incidents on urban streets, accidents happening at intersections could be the most serious; and their impact can be catastrophic and trigger gridlock on the local scope of traffic network. Thus, a key strategy for reducing resulting loss is to handle accidents and incidents as quickly as possible to keep traffic flowing and improve the safety of victims. It is clear that an accident or an incident has to be detected and verified before any other incident management actions can be taken to guarantee the success of any incident management process. Timely, accurate accident detection at intersections therefore becomes more important in any Traffic Management System.

Though various Incident Management Systems have been developed in many places of the United States, most of them mainly provide direct measurement for counting, occupancy measurement, presence detection, queue detection, speed estimation, and vehicle classification. However, in real-time traffic accident detection, the conventional incident management system is less effective. They can only provide inferred detection of incidents by reacting to the symptoms of incidents rather than to the incidents directly. Furthermore, under adverse weather and bad lighting conditions, they often produce high false alarm rates. To overcome these shortcomings, Automated Accident Detection at Intersections system was designed to use acoustic signals to detect traffic accidents at intersections. The system's design idea was motivated by the fact that traffic accidents have characteristics that make them distinguishable from the normal traffic background events. In digital signal processing, the wavelet analysis technique's excellent localization in time and frequency is able to capture the very short-term audio pattern of the accident. The extracted features provide sufficient information for an automated algorithm to detect an accident from normal traffic background events. The audio sensor equipment used in the system is cost-effective and needs little maintenance and management.

The system design strategy simulates the function of the human hearing system, which uses low-level receptors for early stage hearing processing and high-level cognition for pattern recognition and acoustic signal understanding. This model was specially developed to classify acoustic signals into crash, non-crash, or other categories. The system architecture is described as follows. The input to the system is a 3 second segment of audio signal. The system can be operated in two modes: two-class and multi-class. The output of the two-class mode is a label

of “crash” or “non-crash”. In the multi-class mode of operation, the system identifies crashes as well as several types of non-crash incidents, including normal traffic and construction sounds. The system is composed of three main signal processing stages: feature extraction, feature reduction, and classification. Five methods of feature extraction are investigated and compared; these are based on the discrete wavelet transform, fast Fourier transform, discrete cosine transform, real cepstral transform, and mel frequency cepstral transform. Statistical methods are used for feature optimization and classification. Three types of classifiers are investigated and compared; the nearest mean, maximum likelihood, and nearest neighbor methods.

Field data collected from the intersections of Jackson, MS, and Starkville, MS, was used to train and test this detection system. The lab testing results of the study show that the wavelet-based features in combination with the maximum likelihood classifier form the optimum design. The system is computationally inexpensive relative to the other methods investigated, and the system consistently results in accident detection accuracies of 95% to 100% when the audio signal has a signal-to-noise-ratio of at least 0 decibels. The testing accuracies show that the method is capable of effectively performing crash and non-crash classification of acoustic signals and can meet the requirement of real-time accident detection.

With promising results being achieved in a lab environment, a successfully developed real-time implementation system can automatically detect accidents at intersections, which will greatly enhance the safety and efficiency of the surface street networks. By shortening detection and notification time of the accident, along with reduced accident response time due to accurate accident information, the system can reduce the accident clearance significantly and therefore reduce congestion and delay. The Automated Accident Detection at Intersection system can become an integral component of an Advanced Traffic Management System (ATMS).

CHAPTER 1. INTRODUCTION

1.1 Problem Statement

As the development of cities and rural areas continues to increase, the resulting traffic demand is placing an increasing strain on an already congested traffic network. According to the 1999 Urban Mobility Report of the Texas Transportation Institute, traffic congestion cost the U.S public in 68 urban areas 4.3 billion hours of delay, 6.6 billion gallons of wasted fuel consumed and \$72 billion of time and fuel cost in 1997. The traffic delay due to accidents, breakdown and other incidents accounted for about 57 percent of delay [1]. According to the 2002 Urban Mobility Report of the Texas Transportation Institute, congestion is growing in areas of every size, and congestion costs are increasing in every aspect [2]. According to the data (*Lindley* 1989), by the year 2005, the percentage of congestion due to incidents is expected to increase to 70%. The traffic congestion caused by incidents will cost the nation over \$75 billion in lost productivity. From the above statistical data, it is clear that traffic incidents are a major cause of congestion on the nation's traffic network.

Incidents are any non-recurring events that cause congestions by temporarily increasing demand or reducing the capacity of the traffic network. At the same time the considerable congestions can lead to traffic delay, more fuel consumption, bad air quality, and even secondary accidents. The common types of incidents include accidents, construction and maintenance activities, bad weather, and structural failures. Among the

common types of incidents, accident is the most serious kind, especially fatal accidents, accidents with hazardous materials and accidents with spilled cargo. They can reduce the capacity of roadway to over 85%, even 100% [3], of the normal capacity.

On urban surface streets, intersections are the most complicated and dangerous locations within the traffic network. At intersections, vehicular flows from several different approaches making left-turn, through, and right-turn movements seek to occupy the same physical space. The majority of traffic accidents occur at or near intersections, where the resulting loss is the most serious. Their impacts can be catastrophic and trigger gridlock on the local scope of a traffic network [3]. Thus, a key strategy for reducing delay in major urban areas is to handle accidents and incidents as quickly as possible to keep traffic flowing. It is clear that an accident or an incident has to be detected and verified before any other incident management actions can be taken to guarantee the success of any incident management processes. Because accidents happening at intersections can cause the most serious loss of any kind of incident, timely and accurate accident detection at intersections becomes more important in incident management.

While freeway incident detection research can be dated back to the early 1960's, for example, in 1961 Chicago started the Minutemen Program to deal with incident response [4], and California algorithms were developed in 1960s to detect freeway incidents [5]. Accident detection research on surface streets has been lagging and needs more attention. A successfully developed system that automatically detects accidents at intersections in real time can become an integral component of an Advanced Traffic Management System

(ATMS) and enhance the safety and efficiency of the surface street networks. Quick and accurate detection of accidents at intersections is essential so that the necessary medical and emergency response can be provided in the most timely manner possible. By shortening detection and notification time of the accident, along with reduced incident response time due to accurate incident information, the system can reduce the accident clearance significantly and therefore reduce congestion and delay.

1.2 Background

Detecting real-time traffic accidents is a challenging task in an Advanced Traffic Management System. In recent years, technological innovations have given rise to many new types of advanced traffic sensors. A large number of detection systems have been implemented to monitor traffic conditions, including magnetic, ultrasonic, microwave, infrared light, and optical beam sensors. Most of them mainly provide direct measurement for counting, occupancy measurement, presence detection, queue detection, speed estimation, and vehicle classification [6]. However, in real-time traffic accident detection, the conventional incident management system has encountered three basic problems that are difficult to overcome.

First of all, the automatic incident detection algorithm can be divided into two general categories: those based upon pattern recognition and those based upon short-term statistical forecasting. These two kinds of detection techniques perform accident detection based upon a single comparison between observed traffic flow parameters with those

expected, using a pre-determined pattern in historical data [7]. In this detection algorithm, input vectors are made up of the traffic flow parameters such as volume, occupancy, speed, etc instead of the direct characteristics of traffic accidents. That is, this technique can only provide inferred detection of accidents by the symptoms of the accidents, rather than by the accidents directly. On the other hand, to some degree these algorithms are based on preset thresholds calibrated from "normal flow" conditions. When the conditions change, the algorithm's performance will likely degrade on one or more of the measures of performance. Traffic accidents can be caused by many factors such as weather conditions, traffic conditions, vehicle drivers, and other undecided factors. It is very complicated to calibrate the appropriate thresholds. So the automatic incident detection algorithms in conventional incident detection system are less effective for traffic accident detection. Second, many conventional sensors have high false alarm rates. In incident management system the inductive loop detector is the most common sensor, which has been historically prone to high failure rates. The reason for this failure is due to the disturbance caused by the traffic incidents in low flows or due to the shoulder incidents that are so light that these incidents are difficult to be detected [8]. Another defect of the conventional sensor system is high weather sensitivity. Some sensor systems like video cameras suffer an increase in false alarm rates under adverse environments and bad lighting conditions including darkness, precipitation, fog, or dust. Furthermore, their implementation tends to be quite labor intensive and needs much money on purchasing the devices and for their installation, maintenance, and management.

1.3 Project Review

Given these shortcomings of the conventional sensor system that are unavoidable, a new accident detection system was designed using the audio signal to implement automated accident detection at intersections. The system's design idea was motivated by the following conditions:

- Traffic accidents have characteristics that distinguish them from the normal traffic background events and are immediately recognizable to human ear (screech of tires, breaking glass, and huge crash sound). These characteristics are represented in the very short-term audio pattern of the accident. They provide sufficient information for an automated algorithm to detect accidents from normal traffic background events (vehicle passing, vehicle braking, vehicle sirens, construction noise, and some thunder).
- Audio sensor, such as simple microphones, can be cost effective, computationally efficient alternative and needs little maintenance and management. Furthermore, the audio sensor offers some advantages over other sensor systems in the adaptability of environments. It works equally well in all lighting conditions, wide temperatures, and humidity extremes.
- Wavelet Analysis has been invigorated as an inspiring new technique in digital signal processing. Because of its excellent localization in time and frequency, it has been applied to analyze and process the non-stationary signals, such as accident sounds in a traffic audio signal. Wavelet transform has been used for feature

extraction. The wavelet transform of a signal results in a set of detailed and approximate coefficients. These coefficients provide a means of separating fine scale (very localized) behavior from large-scale (global) behavior in the audio signals.

The project “Automated accident detection at intersections” applies new signal processing algorithms to passive acoustic data to implement the accident detection at intersections. There are two layers in the system architecture. In the lower layer, the Discrete Wavelet Transform is used to extract features of acoustic signals. In the higher layer, the statistical classifier is used to process the extracted feature vectors to perform traffic event classification task. The input to the system is a three second segment of audio signal. The system can be operated in two modes: two-class and multi-class. The output of the two-class mode is a label of “crash” or “non-crash”. In the multi-class mode of operation, the system identifies crashes as well as several types of non-crash incidents, including normal traffic and construction sounds. The system block diagram is shown in Figure 1-1.

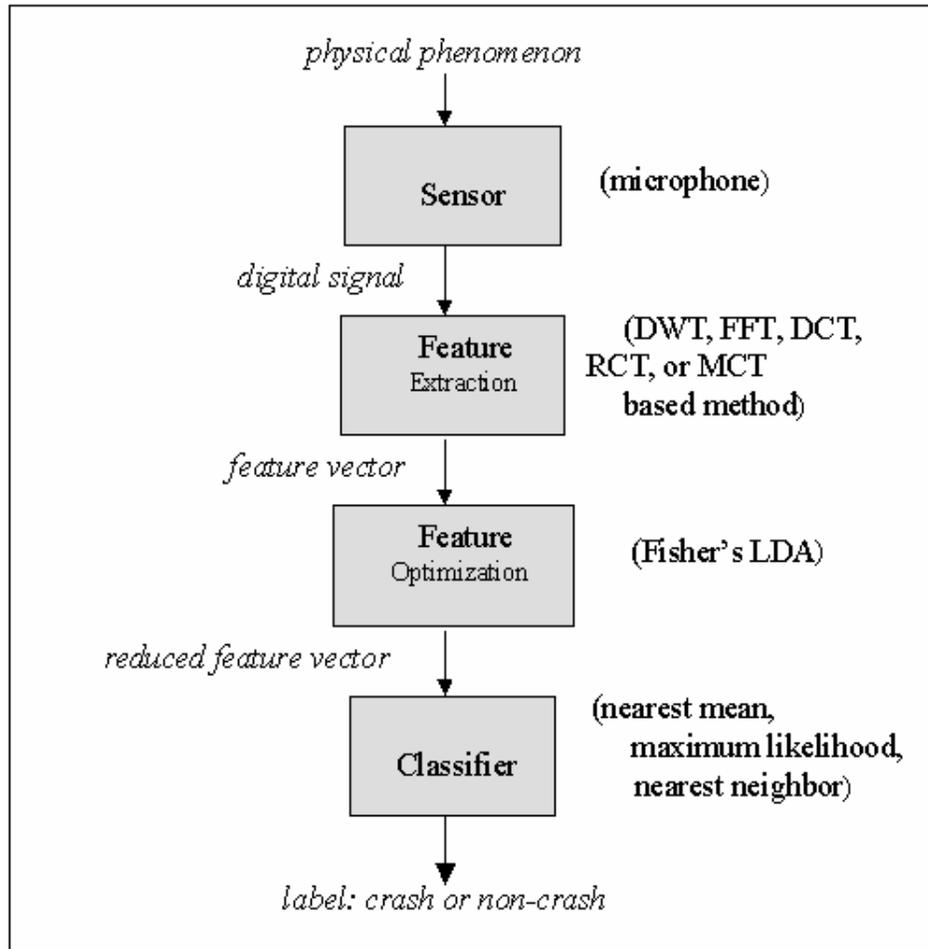


Figure 1- 1 Block diagram of automated accident detection system

To measure the performance of the detection system, the wavelet-based method is compared to various other methods including real cepstral coefficients, mel frequency cepstral coefficients, fast Fourier transform coefficients, and discrete cosine transform coefficients. Furthermore, various classifiers are investigated and compared. These include the nearest mean, maximum likelihood, and nearest neighbor methods. The system was tested on recorded audio signals of normal traffic and traffic accidents. The results showed

that the combination of discrete wavelet transform and maximum likelihood classifier produced the best result.

The “Automated accident detection at intersections” system will perform reliable automatic nearly instantaneous all-weather accident detection, under highly variable traffic conditions. The system can “hear” an accident and make an instantaneous response before congestion builds. The system also overcomes shortcomings of the conventional sensor system, such as high false alarms of loop detectors and high weather sensitivity of video detectors. The keys to the “Automated accident detection at intersections” system are the signal analysis and detection algorithms. The algorithms overcome shortcomings of conventional detection and identification techniques by:

- Taking advantage of wavelet analysis’s excellent localization in time and frequency to capture non-stationary characteristics of audio signals. These characteristics are the most important part of audio signals that could signal an accident.
- Using leave-one-out testing and statistical classifier techniques to classify crash and non-crash audio patterns, accurately identify accidents, and effectively lower false alarm rates.
- Varying the signal-to-noise ratio of the accident data to test detection algorithm’s sensitivity.
- Reporting each decision with a probability-based confidence level to improve decision quality and allow “self-tests” of the system’s performance.

1.4 Project Objective

The objective of Automated Accident Detection system is to facilitate efficient accident detection at major intersections. The achievement of the system will not only save lives and property, but also will minimize the effects of accidents on traffic congestion and reduce the possibility of secondary accidents, thus greatly enhancing the safety and mobility of the transportation system. This can be accomplished by the following:

- Reducing the time spent for accident detection and verification.
- Reducing response time by the appropriate agency.
- Reducing the probability of secondary accidents.
- Reducing the time that personnel are exposed to the accident site.
- Reducing motorist delay.
- Improving travel time reliability.

CHAPTER 2. LITERATURE REVIEW

Most traffic management systems develop various automatic incident detection (AID) methods to detect and respond to traffic incidents as timely as possible. While much research has gone into automatic incident detection on freeways, the complications of the intersections and the special characteristics of traffic accidents have hindered the development of effective automatic accident detection methods at intersections. The literature review focuses upon reviewing the existing researches on automatic incident detection technologies and seeks for an efficient method of automated accident detection at intersections.

The literature review involves three areas that are related to traffic detectors and detection algorithms. The three subject areas are as follows:

1. Traffic detectors.
2. Feature extraction technologies.
3. Feature classification methods.

2.1 Traffic Detectors

Traffic sensor is the cornerstone of any traffic management system. The type of detectors used will determine the kind of traffic data that can be obtained and the accuracy and completeness of those data. In recent years, technological innovations have given rise to many new and different types of advanced traffic detectors that use magnetic, ultrasonic, microwave, infrared light, and optical beam sensors. They provide direct measurement for

presence detection, counting, occupancy measurement, speed estimation, queue detection, vehicle classification, and accident detection.

The “Review of Current and Future Data Requirements and Detector Technologies and the Implications for UTMC” (2002) [9] summarized the most widely used detector technologies in the traffic management environment. These detector technologies were introduced and compared from the application, detection accuracy, and installation/maintenance topics. The main detectors include pressure detector, inductive loop detector, passive acoustic detector, microwave detector, passive infrared detector, video detector, and weather detectors.

2.1.1 Pressure Detector

Pressure detector is often used to measure pulsed presence, count, volume, weight, speed, and pedestrian. The accuracy is within -8% to $+8\%$ in weight measurement, and better than 2% in speed measurement. The advantages of pressure detector are portable or fixed installation and low system cost. But this kind of detector needs to be embedded in the road, and Piezo-electric cables suffer from moisture ingress and temperature change.

2.1.2 Inductive Loop Detector

Inductive loop detector is often used to measure true presence, count, volume, speed, and headway. The accuracy is about $\pm 0.5\%$ in count measurement and $\pm 2\%$ in speed measurement. The advantages of inductive loop detector are accurate traffic data, low cost,

and well-defined detection zone. But this kind of detector needs to be embedded in the road, and is sensitive to the installation process and road works.

2.1.3 Passive Acoustic Detector

Passive acoustic detector is often used to measure true presence, count, volume, speed, and headway. The accuracy is about $\pm 6\%$ in count measurement. This kind of detector is easy to install and maintain, further it has low system cost and better environment operation range.

2.1.4 Microwave Detector

Microwave detector is often used to measure pulsed presence, count, volume, direct speed, and headway. The accuracy is within $\pm 2\%$ in measurement of count. This kind of detector detects velocity directly. It does not disturb traffic during installation and is not affected by the environment. But it does not work well when speed measurement is below 8kph and stationary vehicles are detected.

2.1.5 Passive Infrared Detector

Passive infrared detector is often used to measure true presence, count, volume, speed, and headway. The accuracy is with $\pm 3\%$ in count measurement. It can operate day and night and can be installed overhead and side. But it is sensitive to bad weather conditions other than high temperature exhausts.

2.1.6 Video Detector

Video detector is often used to measure true presence, count, volume, speed, and incident detection. Video detectors can detect large number of parameters. The CCTV system is relatively low cost. But its performance can be affected by snow and rain. In some street applications, the cost is very high.

2.1.7 Weather Detector

Weather detector is often used to measure wind speed, wind direction, and temperature mainly. In general this kind of detector needs high cost.

2.2 Feature Extraction Technology

Traffic accidents have the characteristics that distinguish them from the normal traffic background events and are immediately recognizable to human ears. These characteristics are represented in the very short-term audio pattern of the accident. They provide sufficient information for an automated algorithm to detect an accident from average traffic background events. Digital signal processing techniques are applied to extract features. In this section, five feature extraction methods are surveyed.

- Discrete Wavelet Transform (DWT)
- Fast Fourier Transform (FFT)
- Discrete Cosine Transform (DCT)
- Real cepstral transform (RCT)
- Mel frequency cepstral transform (MCT)

2.2.1 Wavelet Analysis and DWT

Wavelet analysis is based on the idea of projecting a signal onto a set of basis functions. A set of wavelet basis functions, $\{\mathbf{y}_{a,b}(t)\}$ can be generated by shifting and scaling the basic or mother wavelet, $\mathbf{y}(t)$ according to the following,

$$\mathbf{y}_{a,b}(t) = \frac{1}{\sqrt{a}} \mathbf{y}\left(\frac{t-b}{a}\right) \quad (2-1)$$

where $a > 0$ and b are real numbers. The variable a is the scaling factor of a particular basis function, and b is the translation variable along the function's range. When $a > 1$, the functions are dilated and when $a < 1$, the functions are contracted. The coefficient, $\frac{1}{\sqrt{a}}$, is included to normalize the energy of the wavelets. All of the wavelets $\{\mathbf{y}_{a,b}(t)\}$ generated by shifting and scaling the mother wavelet, $\mathbf{y}(t)$, have the same basic shape.

All wavelet functions must oscillate, have an average value of zero, and have finite support. This "admissibility condition" can be represented by

$$\int_{-\infty}^{+\infty} \frac{|\mathfrak{F}(\mathbf{y}(t))|^2}{|s|} ds = 0 \quad (2-2)$$

where $\mathfrak{F}(\cdot)$ denotes the Fourier transform and s is the Fourier domain variable. An important property of many wavelet systems is the multiresolution analysis (MRA) property, where the decomposition of a signal is in terms of the resolution of detail signals [10]. If a wavelet basis satisfies the MRA criteria, its transform can be implemented via multiresolutional filter trees. This type of implementation is very useful because it allows for fast algorithms similar to the well-known fast Fourier transform.

There exist many different types of mother wavelets and wavelet bases. The Haar wavelet is one of the simplest examples. The Haar wavelet is discontinuous, and it resembles a step function:

$$\mathbf{y}(t) = \begin{cases} 1, & 0 \leq t \leq 1/2 \\ -1, & 1/2 \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (2-3)$$

Also, a well-known family of wavelets was developed by Ingrid Daubechies [11], and they are generally referred to as Daubechies- n , where n is the “order” of the mother wavelet. The order corresponds to the regularity of the mother wavelet, and the Daubechies-1 wavelet is equivalent to the Haar wavelet. For this study, the authors investigated the use of many different mother wavelets, including the Coiflet, symlet, biorthogonal, and Daubechies families.

The CWT, denoted by $W_f(a, b)$, of a function, $f(t)$, with respect to the wavelet basis function, $\mathbf{y}_{a,b}(t)$, can be defined as

$$W_f(a, b) = \int_{-\infty}^{\infty} f(t) \mathbf{y}_{a,b}(t) dt \quad (2-4)$$

where the wavelet function $\mathbf{y}_{a,b}(t)$ is given by equation (2-1). For the CWT, the scale parameter, a , and the shift parameter, b , are specified as real numbers. Hence, the transform coefficients, $W_f(a, b)$, are continuous with respect to the variables a and b . For the DWT, the discrete wavelet basis functions are represented as

$$\mathbf{y}_{j,k}(n) = 2^{-\frac{j}{2}} \mathbf{y}(2^{-j}n - k) \quad (2-5)$$

and the wavelet coefficients are obtained by

$$W_{j,k} = \langle f(n), \mathbf{y}_{j,k}(n) \rangle \quad (2-6)$$

Thus, the scales are $a = 2, 4, 8, \dots, 2^j, \dots, 2^p$. Note that the audio signal and wavelets are now functions of discrete time, n , rather than continuous time, t . In this project, $f(n)$ is the digitized audio signal of traffic noises.

The DWT has been extensively used in the development of fast wavelet algorithms. The most common implementation of the DWT is the well-known dyadic filter tree [12]. The filter tree is composed of high-pass and low-pass filters corresponding to the user's selection of mother wavelet function. At each stage j of the filter tree, a set of approximation, A_j , and detail coefficients, D_j , are produced, corresponding to the input signal's large and small-scale behavior, respectively.

2.2.2 Fast Fourier Transform (FFT)

Fast Fourier Transform (FFT) is a computationally efficient version of the Discrete Fourier Transform (DFT) [13]. It is based on divide-and-conquer approach in which the DFT is divided into smaller and simpler problems and the final DFT is rebuilt from the simpler DFTs. When the time data and their DFT are too large to be stored in the main memory, the FFT is done in parts and the results are pieced together to form the overall FFT, and saved on secondary storage devices such as hard disk.

In the simplest Cooley-Tukey version of the FFT, the dimension of the DFT is successively divided in half until it becomes unity. This requires the initial dimension N to be a power of two. The FFT algorithm is defined by the equation below.

$$X(k) = \sum_{n=0}^{N/2-1} W_N^{k(2n)} x(2n) + \sum_{n=0}^{N/2-1} W_N^{k(2n+1)} x(2n+1) \quad k = 0, 1, \dots, N-1 \quad (2-7)$$

Where $W_N = e^{-j2\pi/N}$ and $W_N^{kn} = e^{-j2\pi kn/N}$

The typical FFT algorithm consists of three conceptual parts:

1. Shuffling the N-dimensional input into N one-dimensional signals.
2. Performing N one-point DFTs.
3. Merging the N one-point DFTs into one N-point DFT.

2.2.3 Discrete Cosine Transform (DCT)

The discrete Cosine Transform (DCT) [14] is a technique that converts a spatial domain waveform into its constituent frequency components as represented by a set of coefficients. Typically the DCT coefficients produced have most of the block's energy in a few frequency domain elements.

The 1-D DCT algorithm is defined by the equation below.

$$X(k) = \mathbf{a}(k) \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\mathbf{p}(2n+1)k}{2N}\right] \quad 0 \leq k \leq N-1 \quad (2-8)$$

where

$$\mathbf{a}(0) = \sqrt{\frac{1}{N}}, \quad \mathbf{a}(k) = \sqrt{\frac{2}{N}} \quad \text{for } 1 \leq k \leq N-1 \quad (2-9)$$

The features of DCT are as follows:

- DCT is orthonormal
- Real outputs respond to real inputs

- DCT has better energy compaction (much of the signal energy can be represented by only a few coefficients)
- Its coefficients are nearly uncorrelated
- DCT has more than one version

2.2.4 Real Cepstral Transform (RCT)

The real cepstrum [15] of a digital signal $x[n]$ is defined as

$$c[n] = \frac{1}{2n} \int_{-\pi}^{\pi} \ln |X(e^{jw})| e^{jwn} dw \quad (2-10)$$

Cepstral transformation is based on homomorphic transformation that separates the source from the filter, where logarithms are used to convert multiplicative terms into additive terms.

Typically in speech processing, the first 12 to 20 RCT coefficients are utilized.

2.2.5 Mel Frequency Cepstral Transform (MCT)

Mel frequency cepstral transform [15] is a real cepstral transform of a windowed short-time signal, which is derived from the fast Fourier transform of the signal. The basic difference between RCT and MCT is that a non-linear frequency scale is used in MCT, which basically approximates the behavior of the system. Let the DFT of given input signal be

$$X_a[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/N}, \quad 0 \leq k \leq N \quad (2-11)$$

and also a filter bank with M filters is defined ($m=1,2,\dots,M$), where filter m is triangular filter which is given by:

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{(k - f[m-1])}{f[m] - f[m-1]} & f[m-1] \leq k \leq f[m] \\ \frac{(f[m-1] - k)}{(f[m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (2-12)$$

which satisfies $\sum_{m=1}^M H_m[k] = 1$.

Let the lowest and the highest frequencies of the filter bank be f_l and f_h respectively, F_s be the sampling frequency in Hz, M be the number of filters, and N be the size of the FFT. The boundary points $f[m]$ are uniformly spaced in the mel-scale:

$$f[m] = \left(\frac{N}{F_s} \right) B^{-1} \left(B(f_l) + m \frac{B(f_h) - B(f_l)}{M+1} \right) \quad (2-13)$$

where the mel-scale B is given by $B(f) = 1125 \ln(1 + f/700)$ and

$$B^{-1}(b) = 700(e^{(b/1125)} - 1) \quad (2-14)$$

The log-energy at the output of each filter is

$$S[m] = \ln \left[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k] \right], \quad 0 < m \leq M \quad (2-15)$$

The mel-frequency cepstrum is then the discrete cosine transform of the M filter outputs:

$$c[n] = \sum_{m=0}^{M-1} S[m] \cos(\Pi n(m-1/2)/M) \quad 0 \leq n < M \quad (2-16)$$

where M varies for different implementations from 24 to 40. For speech recognition, typically only the first 13 cepstrum coefficients are used.

2.3 Feature Classification Methods

Three types of statistical classifiers are investigated: nearest mean, maximum likelihood, and nearest neighbor [16]. All three are supervised classifiers. That is, the classifier must first be trained using data for which the user knows the correct classification.

2.3.1 Maximum Likelihood Classifier

The maximum likelihood classifier is a parametric classifier that requires second order statistics of the training data, *i.e.* class means and variances. Using the class statistics of the training data, class boundaries are computed. The test data is then compared to the class boundaries. Whichever class's boundaries encompass the test data is then selected as the class of the test data. It is more complicated to implement than the nearest mean, however, if the data is not symmetrically distributed, the maximum likelihood method will outperform the nearest mean classifier.

2.3.2 Nearest Mean Classifier

It is a parametric classifier that requires only first order statistics of the training data, *i.e.* class means. Using the training data, the mean of each class is computed. Then the test data is compared to the class means. The Euclidean distance between the test data and each class mean is computed. The class with the shortest distance is chosen as the class of the test data.

2.3.3 Nearest Neighbor Classifier

The nearest neighbor classifier is non-parametric. The test data is compared to all of the training data. The Euclidean distance is computed between the test data and each sample in the training data. The test data is then assigned to the same class as that of the training data with the shortest distance (the nearest neighbor to the test data). Note that with the nearest mean and the maximum likelihood classifiers, only the class statistics (means and variances) are stored for use during the testing phase. On the contrary, all training data must be stored for use during the testing phase of the nearest neighbor classifier, and this causes the system to need significantly more memory and computational power. However, the nearest neighbor method can significantly outperform the others when the classes are not unimodal or possess a significant number of outliers.

CHAPTER 3. RESEARCH TASK AND PROCEDURE

3.1 Data Collection and Processing

In developing the accident detection system, the first step is to establish the acoustic database, which provides the fundamental resource for detection algorithm. The acoustic database contains two types of acoustic data: background traffic sounds and traffic accident sounds. Because these acoustic data are mainly used to train detection algorithm, the data input to these detection algorithms must be known.

3.1.1 Ordering Recording Equipment

The background traffic sounds in the acoustic database were obtained by on-site collection using recording equipment. Before ordering the equipment, a comprehensive review of the range of detectors was carried out, including the parameters they can measure, the communication links required, the installation and maintenance, and an assessment of their advantages and disadvantages. Based on these estimations, a set of recording equipment was ordered. The recording equipment list is as follows:

- Sony TCD-D8 DAT walkman –Digital Audio Tape Recorder
- Super Mono microphone (SMM –1) Omni-Directional
- DAT cassette
- External sound card
- S/PDIF Coaxial cable

3.1.2 Data Collection

Two approaches were taken to generate an acoustic database. Traffic background sounds were collected at intersections in Starkville, MS, and Jackson, MS. Traffic crash sounds were obtained from the crash test facility.

The background traffic sounds in the acoustic database were obtained by on-site collection using Sony TCD-D8 Digital Audio Tape Recorder and Super Mono microphone (SMM –1). The sound data were recorded at the sample frequency of 44.1 Khz. The field recordings at the intersections of Jackson and Starkville were digitized for use in generating the initial acoustic database of traffic background noise. To complete the database, data were respectively collected under dry and humid daylight conditions. Normal acoustic signals are from various vehicles and traffic conditions, including sounds produced by trucks, cars, motorcycles, and buses in normal operation, brake sounds, vehicle horn sounds and the sounds from construction and industrial activity near the intersections. Data collections were conducted in the city of Jackson at locations recommended by the MDOT, and at several busy intersections in the city of Starkville. The sounds were collected under different traffic conditions and on different weekdays. In addition, other sound samples were gathered from web sites for potential use. The collection locations in the city of Jackson are as follows:

- Intersection of Highway 80@Ellis AVE,
- Intersection of Woodrow Wilson AVE@State Street,
- Intersection of Count Line RD@US 51,
- Intersection of Northside Dr@I-55 Frontage,

- Intersection of Highway 25/Lakeland@I-55 Frontage,
- Intersection of Highway 25/Lakeland@Ridgewood,
- Intersection of Highway25/Lakeland@Hwy 475/Airport Rd
- Intersection of Hwy 80@Hwy 49/Hwy 468.

The traffic sounds were collected at the intersection of Hwy 25/Lakeland@Hwy 475/Airport Rd and at the intersection of Count Line RD@US 51 under humid conditions.

The collection locations in the city of Starkville are as follows:

- Intersection of Jackson Street@main Street.
- Intersection of Highway 12@Highway 25.
- Intersection of Highway 12@Montgomery Street.

Acoustic data of crash events are difficult to collect in the real world. Charles Harlow of Louisiana State University who undertook a similar study provided the crash audio data, collected at the Texas Transportation Institute Crash Test Facility. In addition, the data provided by Charles Harlow also contains routine traffic sounds from a construction site, and the braking sounds from severe braking with no crash. In order to mimic practical scenarios, the audio data collected from the test center were processed before they were used.

3.1.3 Data Description and Preprocessing

The digital recordings of the two kinds of sounds were converted into a common ‘WAV’ format that can be read by both Matlab and a standard sound card on the PC. The sampling specifications were 22050 HZ sampling rate, 8-bit resolution and mono-channel. The signals were partitioned using a 3 second rectangular window, and this resulted in each

audio signal having a length of 66176 samples. The detection algorithm developed using the Matlab software can access the data in a common format.

In some audio data, especially the crash audio signal, the instantaneous power varies too much to get a good measure of the original signal's maximum sustained power, which is likely to be the main characteristic of the signal. In order to get an estimation of the instantaneous power over time, before any processing was conducted all signals were normalized so that each signal's maximum amplitude was one. Thus normalized signals always fall within a specified range to keep audio signals at a reasonable level.

Acoustic data of crash events are difficult to collect. The crash sound effects were obtained from Texas Transportation Institute Crash Test Facility. In real world, the crash sound varies from the sound obtained from the crash testing centers. In order to mimic practical scenarios, a realistic crash sound was obtained by synthesizing the pure crash signals (obtained from the crash testing centers) with the other available normal background traffic signals. The synthesized crash signal is a weighted sum of a crash signal and a non-crash background signal. Assume $f_{nc}(n)$ is the recorded non-crash signal, $f_c(n)$ is the recorded crash signal, and \mathbf{a} the weighting variable. Then the new, realistic crash signal, $\tilde{f}_c(n)$, was computed as

$$\tilde{f}_c(n) = f_{nc}(n) + \mathbf{a} \cdot f_c(n) \quad (3-1)$$

The authors investigated the automated system's sensitivity to crash proximity. It was assumed that crashes occurring near the microphone (directly at the intersection) would be louder in volume than crashes occurring farther away (in a street entering the

intersection). To model the various scenarios, the weight \mathbf{a} was varied. The resulting crash signal has a signal-to-noise-ratio (SNR) ranging from -50 decibels (dB) to $+50$ dB. When computing the SNR of $\tilde{f}_c(n)$, the “signal” component was $f_{nc}(n)$ and the “noise” component was $f_c(n)$.

The amplitude signatures for pure crash sound and non-crash sound with braking sound are shown in figures 3-1 through 3-2. Signal-to-noise-ratio (SNR) was used to model the above scenarios. SNR is the ratio of the variance of the accident signal to the background signal over the sampling interval. It measures the relative intensity of the signal relative to other traffic sounds. Figures 3-3 through 3-9 show the scaled signatures of the accident signal, the background, and the synthesized accident signal and background for the various SNR's ranging from -50 decibels (dB) to $+50$ dB (50 dB, 20 dB, 10 Db, 0 dB, -10 dB, -20 dB, -50 dB).

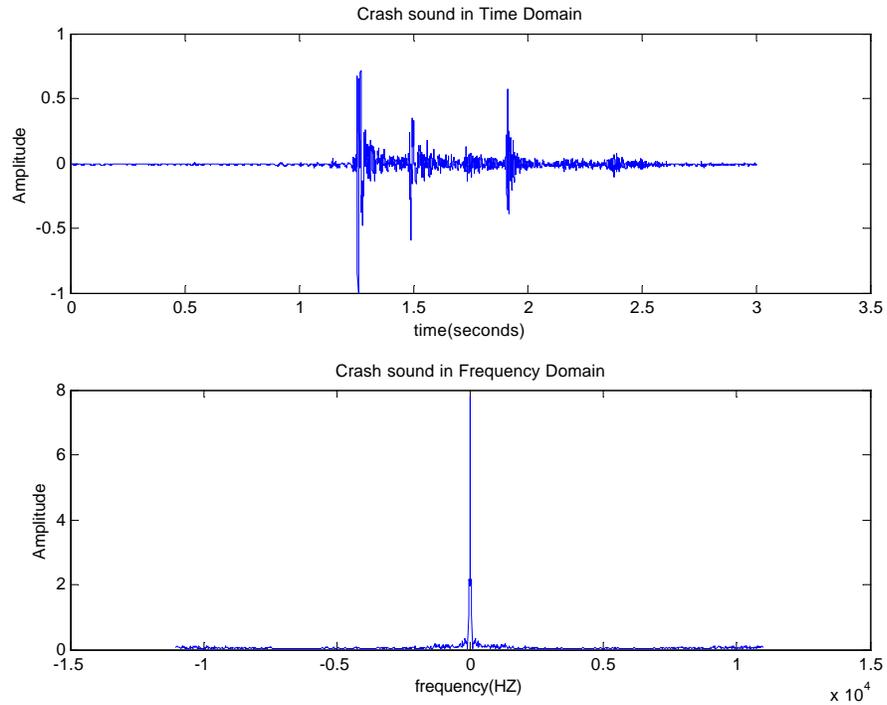


Figure 3- 1 Crash sound in time domain and frequency domain

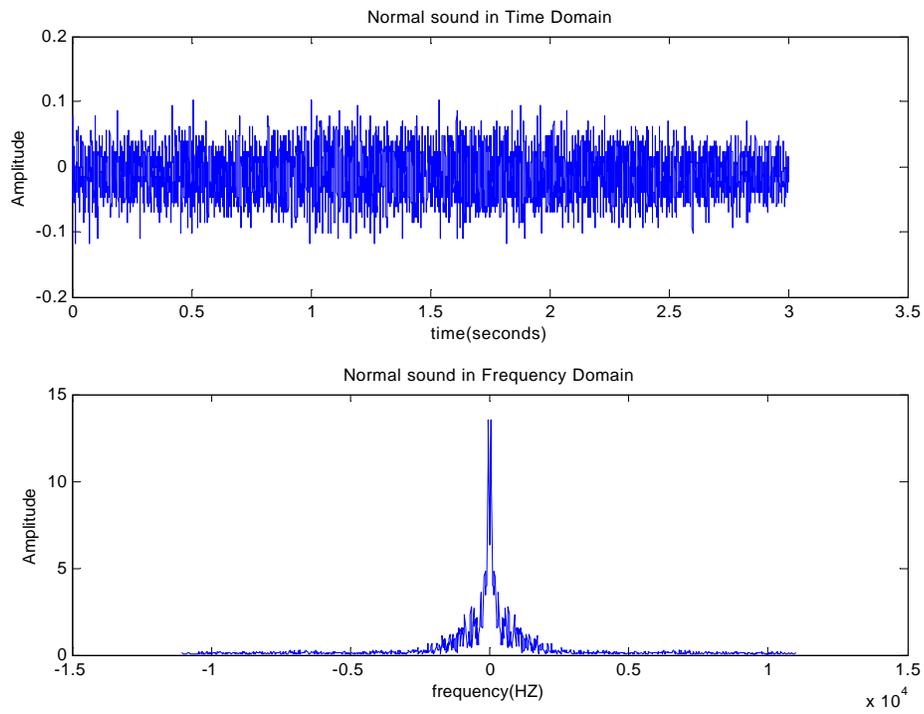


Figure 3- 2 Normal sound in time domain and frequency domain

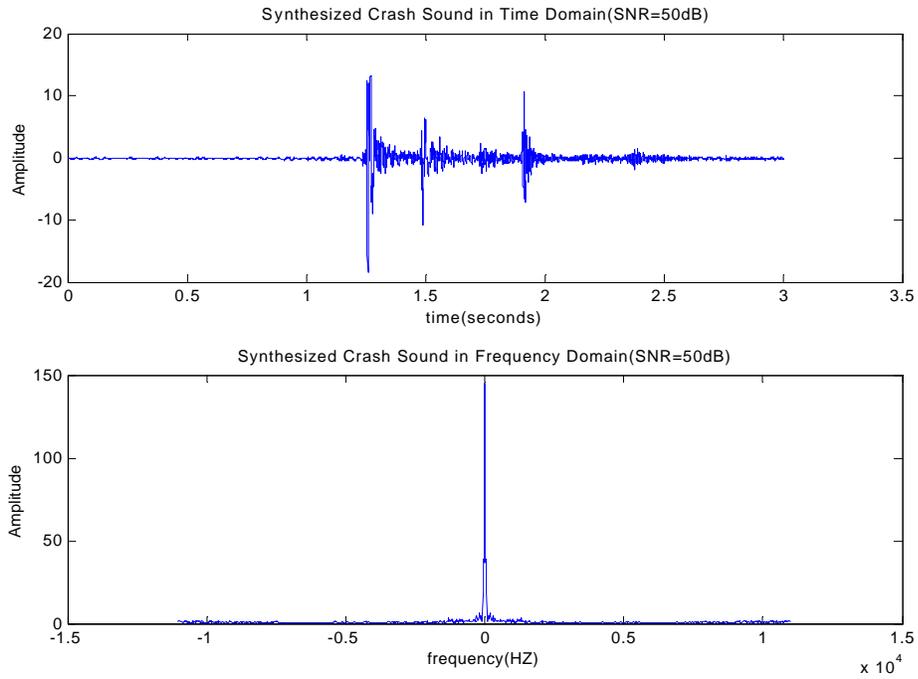


Figure 3- 3 Synthesized crash sound (SNR=50dB)

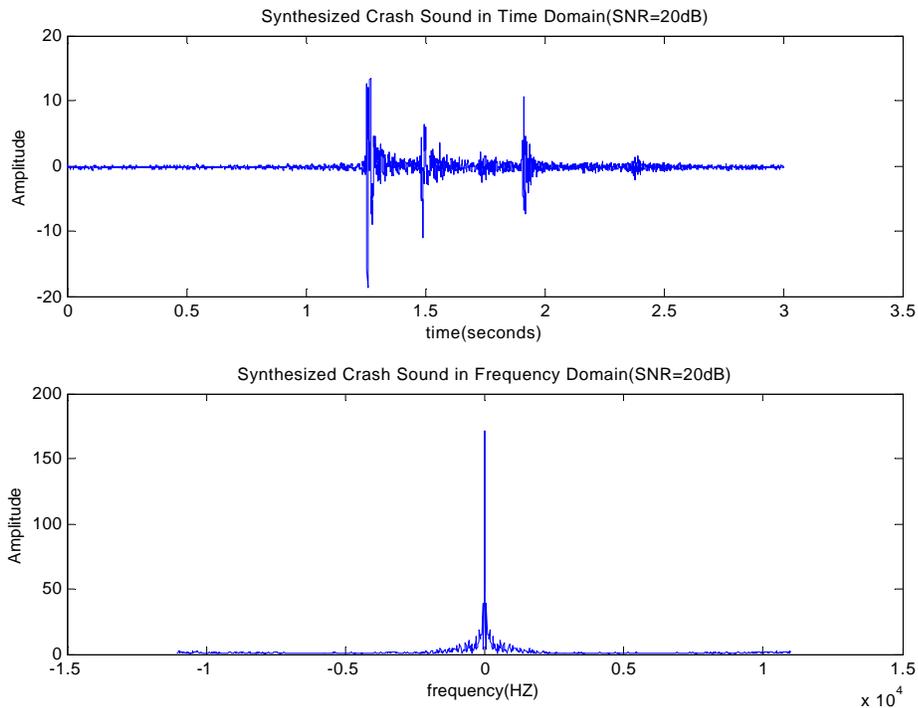


Figure 3- 4 Synthesized crash sound (SNR=20dB)

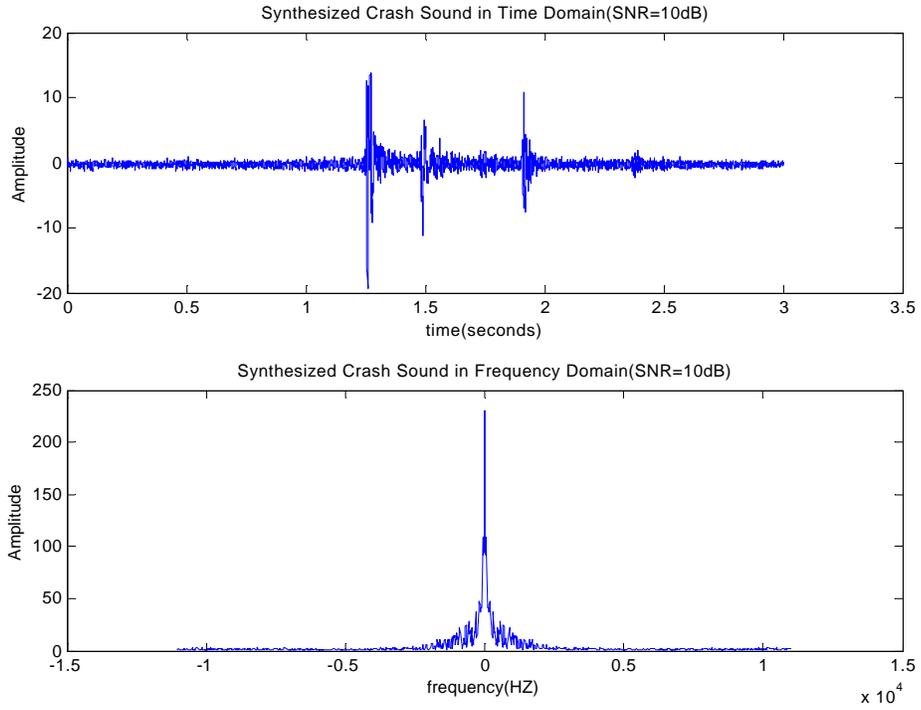


Figure 3- 5 Synthesized crash sound (SNR=10dB)

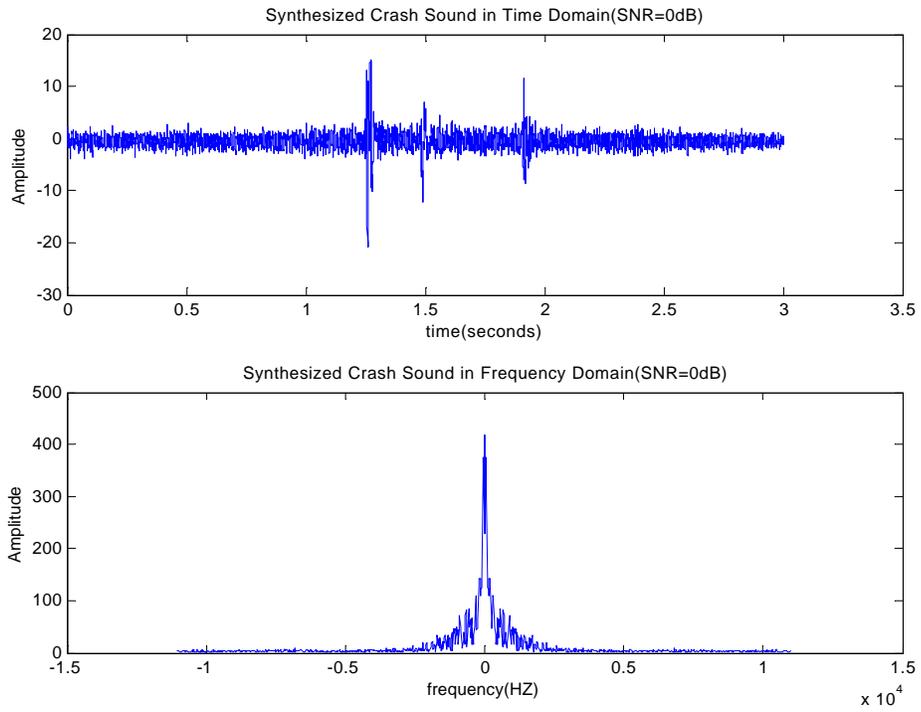


Figure 3- 6 Synthesized crash sound (SNR=0dB)

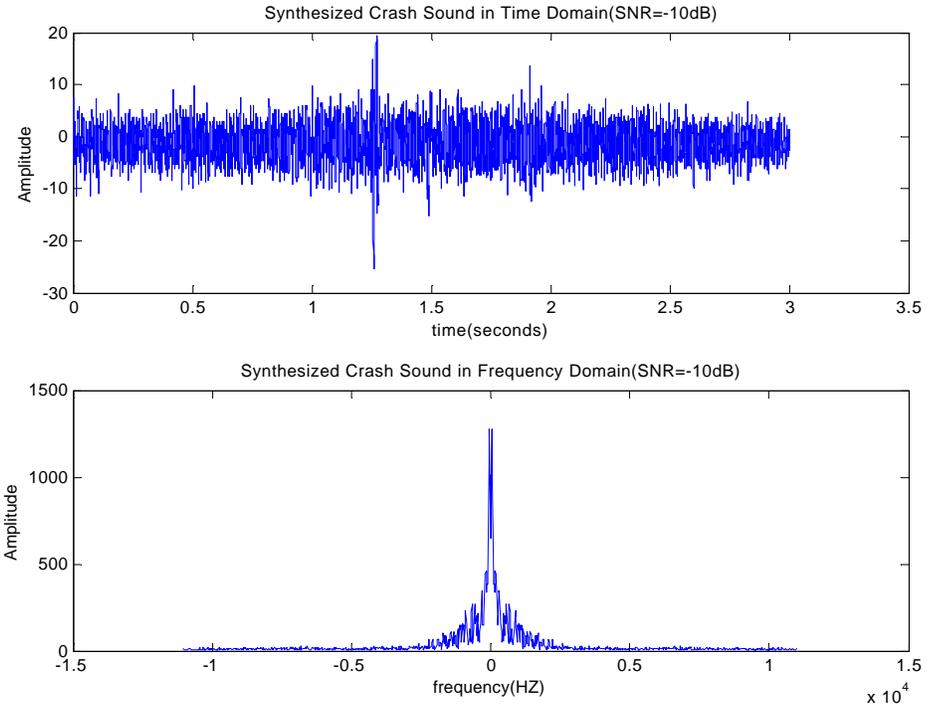


Figure 3- 7 Synthesized crash sound (SNR=-10dB)

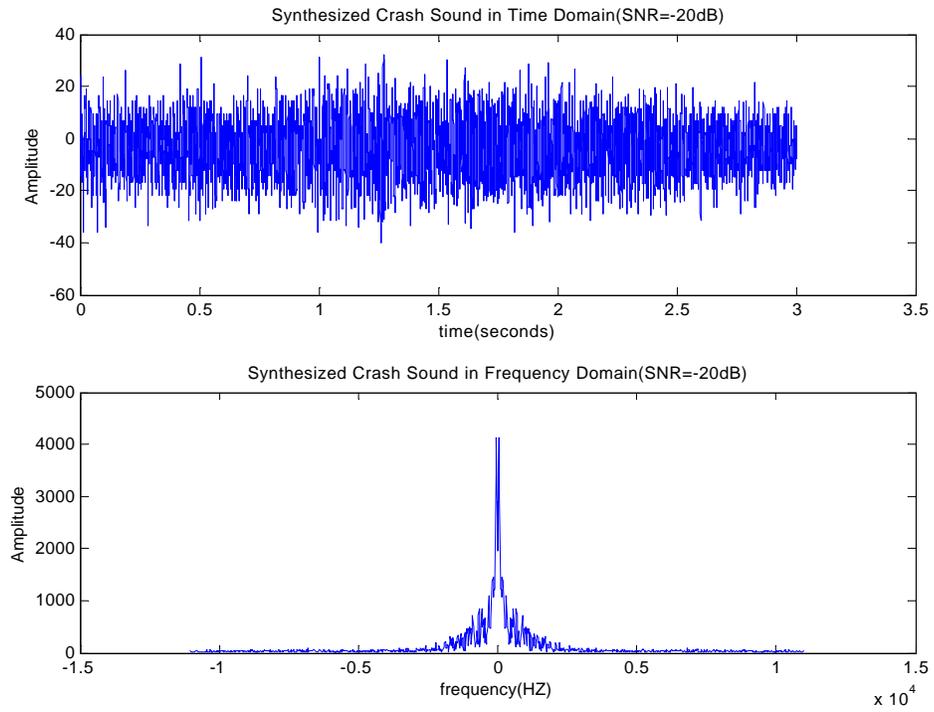


Figure 3- 8 Synthesized crash sound (SNR=-20dB)

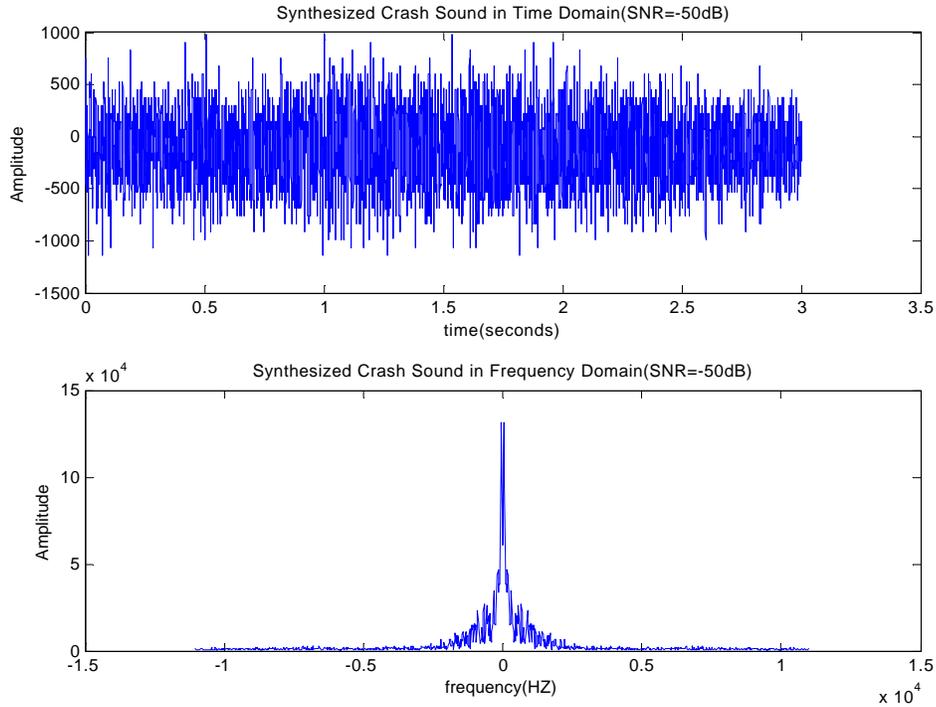


Figure 3- 9 Synthesized crash sound (SNR=-50dB)

3.2 Feature Extraction

Extracting the differing features of crash signal and background signal is an important step in developing accident detection system. The purpose of feature extraction is to characterize the accidents and background acoustic signals by a series of numbers (feature vector). The extracted feature vector could most uniquely represent a given class of signal in the presence of other known classes and then could be used to distinguish an accident from the background.

In this project, various feature extraction methods are investigated and compared.

They are

- Discrete Wavelet Transform (DWT)

- Fast Fourier Transform (FFT)
- Discrete Cosine Transform (DCT)
- Real cepstral transform(RCT)
- Mel frequency cepstral transform(MCT)

In the following part, the five methods will be described respectively.

3.2.1 Extraction Based on DWT

The development of the wavelet feature extractor was motivated by the fact that Discrete Wavelet Transform offers some advantages in analyzing the temporal and spectral properties of non-stationary signals like traffic acoustic signals. A feature vector, \vec{F} , is formed through computing the root-mean-square (RMS) energy of the wavelet detail coefficients at each scale, ED_j and the RMS energy of the wavelet approximation coefficients at the final scale, EA_M ,

$$\vec{F} = [EA_M \quad ED_M \quad ED_{M-1} \quad \dots \quad ED_1]^T \quad (3-2)$$

where the superscript T denotes a vector transpose,

$$EA_M = \frac{1}{P_M^A} \sqrt{\sum_{i=0}^{P_M^A-1} [A_M(i)]^2} \quad (3-3)$$

and

$$ED_j = \frac{1}{P_j^D} \sqrt{\sum_{i=0}^{P_j^D-1} [D_j(i)]^2} \quad (3-4)$$

for $j = 1, 2, 3, \dots, M$, the constant M is the maximum wavelet decomposition level; P_j^D is the number of detail coefficients at level j ; and P_M^A is the number of approximation coefficients at level M . According to equation (3-2), after the DWT feature extraction, the data space dimension is reduced to $M + 1$. The value of M is determined by the wavelet

filter length and the original signal length. The most it can ever be is $M = \log_2(N)$, where N is the length of the original signal and the mother wavelet's high-pass and low-pass filters have length 2, as is the case with the Haar mother wavelet. However, for longer mother wavelets, such as the higher order Daubechies or Biorthogonal mother wavelets, the maximum scale M can be much lower.

By computing the RMS of the DWT coefficients at each scale, the features represent a scalar energy partitioning of the original signal. From a digital signal processing perspective, scale is similar to frequency. However, scale is not equivalent to frequency since the basis functions of the DWT are not necessarily sinusoids. Based on the selection of mother wavelet the scalar energy partitioning may be optimized. However, is it advantageous to simply use frequency analysis of the audio signals? To answer this question, the DWT approach was compared to feature extraction based on the fast Fourier transform (FFT).

3.2.2 Extraction Based on FFT

The second method of feature extraction is to simply use the magnitude of FFT coefficients as features. To fairly compare the DWT-based features with FFT-based features, the order of the FFT was varied. The order of the FFT was selected such that the number of resulting FFT coefficients, and thus features, was equivalent to the number of DWT-based features.

3.2.3 Extraction Based on DCT

The third method of feature extraction is based on the discrete cosine transform (DCT) [17]. The DCT is a real-valued transform, and does not involve complex numbers. The DCT coefficients were used directly as the features. Likewise, the order of the DCT was varied such that the number of resulting DCT coefficients, and thus features, was equivalent to the number of DWT-based features. The DCT was investigated because it is similar to the FFT in that it is used for frequency analysis of the input signal. However, its resulting coefficients are real, whereas the FFT's coefficients are complex.

3.2.4 Extraction Based on RCT

The fourth and fifth methods of feature extraction are based on the cepstral transform [15]. Cepstral analysis is based on homomorphic transformations, where logarithms are used to convert multiplicative terms into additive terms. Cepstral analysis has been utilized extensively in speech recognition, and more recently by Harlow and Yu for automated accident detection using audio signals [18]. Cepstral transforms require windowing of the input audio signal, which is three seconds length for this project. The real cepstral transform (RCT) utilizes Hamming windows to segment the input audio signal into 100msec intervals with 50msec overlap. For each signal segment, the RCT results in a set of ordered coefficients. Typically in speech processing, the first 12 to 20 coefficients are utilized [15]. In this project, the first $(M + 1)$ coefficients are used as features, since the DWT approach results in $(M + 1)$ features. Then for each signal segment, the RCT

coefficients are used as features, and statistical feature reduction and classification are conducted. If a “crash” is detected within at least three of the signal segments, an overall classification of “crash” is assigned to the three-second signal. Note that this requires the RCT, as well as the feature reduction and classification, to be completed 60 times. Consequently, the computational requirement for the cepstral method is much greater than for the DWT, FFT, and DCT methods.

3.2.5 Extraction Based on MCT

The fifth method of feature extraction is based on the mel frequency cepstral transform (MCT) [15]. The MCT uses a nonlinear frequency scale, where the RCT uses a linear frequency scale. The nonlinear scale approximates the behavior of the human auditory system. In speech recognition applications, the MCT has been shown to provide superior classification accuracies over the RCT [15]. As with the RCT, the MCT method requires a windowing of the input audio signal. For the MCT, the three-second audio signal is partitioned into 40 segments. For each segment, $(M + 1)$ of the MCT coefficients are used for classification. The method is repeated 40 times, and an overall classification of “crash” is assigned to the signal if at least three of the segments are classified as “crash”. Like the RCT, the MCT is much more computationally intensive than those of the DWT, FFT, or DCT methods.

3.3 Feature Reduction

Whether the feature extraction is based on DWT, FFT, DCT, RCT, or MCT, the input to the automated statistical classifier is a feature vector. In order to optimize the feature vector, it is reduced using Fisher's linear discriminant analysis (LDA) [16]. The reduced feature vector is then the input to a statistical classifier.

The input to LDA is the audio signal's feature vector, \vec{F} . The output from LDA is an optimal linear combination weight matrix, W , in the sense of maximizing the interclass variance and minimizing the intraclass variance. The weight matrix has a size of $(M+1) \times (C-1)$, where C is the number of classes. With the weight matrix, the reduced feature vector, \vec{F}_r , can be computed as an optimal linear combination of elements in the original feature vector, \vec{F} ,

$$\vec{F}_r = W^T \cdot \vec{F} \quad (3-5)$$

3.4 Feature Classification

Three types of statistical classifiers are investigated: nearest mean, maximum likelihood, and nearest neighbor [16]. All three are supervised classifiers. That is, the classifier must first be trained using data for which the user knows the correct classification. The nearest mean classifier is the simplest of the three. It is a parametric classifier that requires only first order statistics of the training data, *i.e.* class means. The maximum likelihood classifier is a parametric classifier that requires second order statistics of the training data, *i.e.* class means and variances. It is more complicated to implement than the

nearest mean, however, if the data is not symmetrically distributed, the maximum likelihood method will outperform the nearest mean classifier. The nearest neighbor classifier is non-parametric. All training data must be stored for use during the testing phase of the classifier.

3.5 System Testing

The overall system is evaluated using the leave-one-out test. Consider the case that we have a finite number of audio signals in a database and the automated detection system and the classifiers must be trained and tested. In order to have unbiased results, the training and testing data must be mutually exclusive. The leave-one-out approach maximizes the use of the limited database. For each signal in the database, the following test is conducted. Remove the signal under investigation. On the remaining signals, where the classifications are known, compute the features. Use LDA to determine the optimum linear combination and reduce the dimensionality of the feature vectors. Utilize the classifier to compare the test signal with the training signals. Refer to the true class of the test signal, and determine if the automated classification is correct. This process is repeated for each audio signal in the database. Therefore, the percentage of correct classifications is determined as well as a final classification accuracy. Since the test signal is not used in the training, the use of leave-one-out test leads to an unbiased classification accuracy [16]. As compared with jack-knifing, the leave-one-out test makes full use of the test data. This is important when the amount of test data is limited, which is often the case when using crash signals. To further account for

limited test data in this study, a 95% confidence interval [19] was also computed and reported with the classification accuracies.

CHAPTER 4.RESULTS AND DISCUSSION

To create the optimal automated accident detection system, the following design parameters must be determined:

- **Feature extraction method** There are five different types of feature extraction methods investigated and tested in the step of feature extraction, which are DWT, FFT, DCT, RCT, and MCT-based methods.
- **Statistical classifier type** There are three different types of statistical classifiers investigated and tested in the step of pattern classification. They are nearest mean, maximum likelihood, and nearest neighbor.
- **Mother wavelet type** Within the DWT approach, seven types of mother wavelets were tested and compared. They are Haar, Daubechies4, Daubechies15, Coiflet2, Coif lets5, Symlets2, and Symlets8 classes of mother wavelets.
- **Classification mode** The audio signals could be classified using a two-class system (crash or non-crash) or a multi-class system (crash, normal traffic, pile drive, and braking). Both approaches are studied.

The testing results are analyzed and compared to decide the optimal design parameters of the automated accident detection system.

Tables 4-1 and 4-2 show the maximum likelihood classification results for the DWT method when varying the type of mother wavelet. The data used for this experiment was normalized such that all three-second signals in the database have a maximum amplitude of

one. Table 4-1 is for the two-class system, and Table 4-2 is for the multi-class system. Clearly, the DWT approach provides excellent results. The Haar, Daubechies4, Coiflets2, and Symlets8 types of mother wavelets typically performed the best. And this is demonstrated with the results shown in Table 4-1.

Methods used:

- Feature extractor: DWT
- Statistical classifier: Maximum Likelihood Classifier
- Leave-one-out testing
- Data normalized with LDA

| Wavelet | Crash | Non-crash | Overall Accuracy | Confidence Interval (95%) |
|--------------|--------|-----------|------------------|---------------------------|
| Haar | 1 | 1 | 1 | 0 |
| Daubechies4 | 1 | 1 | 1 | 0 |
| Daubechies15 | 1 | 0.9722 | 0.9798 | 0.0232 |
| Coif lets2 | 0.9259 | 1 | 0.9798 | 0.0232 |
| Coif lets5 | 1 | 1 | 1 | 0 |
| Symlets2 | 0.964 | 0.9861 | 0.9798 | 0.0232 |
| Symlets8 | 0.9259 | 1 | 0.9794 | 0.0237 |

Table 4- 1 Maximum likelihood classification accuracies for two-class system using DWT-based features

| Wavelet | Crash | Non-crash | Pile-drive | Overall Accuracy | Confidence Interval (95%) |
|--------------|--------|-----------|------------|------------------|---------------------------|
| Haar | 0.963 | 1 | 0.5 | 0.9394 | 0.0393 |
| Daubechies4 | 0.9259 | 1 | 0.8 | 0.9697 | 0.0283 |
| Daubechies15 | 0.9259 | 0.9839 | 0.8 | 0.9495 | 0.0361 |
| Coif lets2 | 0.9259 | 1 | 0.9 | 0.9697 | 0.0283 |
| Coif lets5 | 0.9259 | 1 | 0.9 | 0.9697 | 0.0283 |
| Symlets2 | 0.963 | 1 | 0.8 | 0.9697 | 0.0283 |
| Symlets8 | 0.963 | 0.9839 | 0.875 | 0.9691 | 0.0288 |

Table 4- 2 Maximum likelihood classification accuracies for multi-class system using DWT-based features

Table 4-3 and Figure 4-1 show the classification results of the DWT (Haar), FFT, DCT, RCT, and MCT methods for both the two-class and the multi-class systems. Again, the data used for this experiment was normalized such that all three-second signals in the database have a maximum amplitude of one. The maximum likelihood classifier was used. Note that the two-class system always outperforms the multi-class system. This result is intuitive since all of the non-crash signals are combined into one category, greatly reducing the chances for misclassification. Also note that the wavelet and cepstral transform methods perform the best, achieving greater than 98% and 94% accuracies for the two-class and multi-class systems, respectively.

Methods used:

- Feature extractor: DWT, FFT, DCT, RCT, MCT
- Statistical classifier: Maximum Likelihood Classifier (with LDA)
- Leave-one-out testing
- Data normalize with LDA (0dB)

| Feature Extractor | Two-class | Multi-class |
|-------------------|-----------|-------------|
| DWT | 1 | 0.9394 |
| RCT | 1 | 0.9897 |
| MCT | 0.9899 | 0.9596 |
| FFT | 0.9091 | 0.8788 |
| DCT | 0.9394 | 0.8485 |

Table 4- 3 Maximum likelihood classification accuracies for two-class and multi-class systems

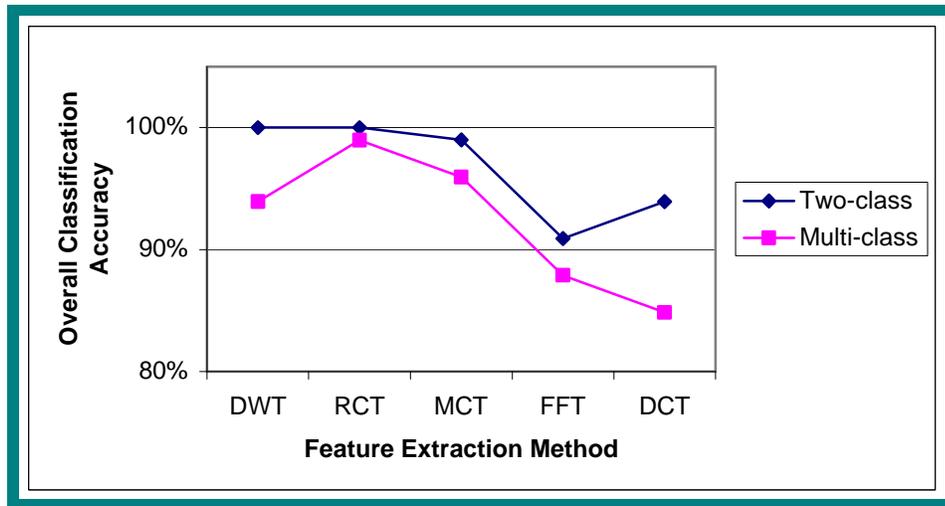


Figure 4- 1 Maximum likelihood classification accuracies for two-class and multi-class systems

Figures 4-2 and 4-3 show the two-class and multi-class classification accuracies of the DWT (Haar) feature extraction method for each of the three types of statistical classifiers. For this experiment, the audio signals were manipulated to model various ambient noise conditions. The SNR's were varied from -50dB to 50dB . Note that the -50dB case represents a scenario that the crash audio signal is very low in amplitude (quiet) as compared with the non-crash audio signals (loud). The 50dB case represents a scenario that the crash audio signal is very high in amplitude (loud) as compared with the non-crash audio signals (quiet). The 0dB case models a scenario that the crash signals and non-crash signals have the same volume. From Figure 4-3, we can see that the classification accuracies decrease with decreasing SNR. This is intuitive since with decreasing SNR, the crash audio signal is becoming more and more like a non-crash audio signal. Note that regardless of SNR, the maximum likelihood classifier performs best.

Methods used:

- Feature extractor: DWT (Haar wavelet)
- Statistical classifier:
 - ✓ Nearest Mean
 - ✓ Maximum Likelihood
 - ✓ Nearest Neighbor
- Leave-one-out testing
- Data normalize with LDA

| SNR | Nearest Mean | Maximum Likelihood | Nearest Neighbor |
|--------|--------------|--------------------|------------------|
| 50db | 0.9697 | 1 | 1 |
| 20 db | 0.9697 | 0.9899 | 1 |
| 10 db | 0.9697 | 0.9899 | 1 |
| 0 db | 0.9596 | 0.9899 | 0.9899 |
| -10 db | 0.9596 | 0.9495 | 0.9596 |
| -20 db | 0.7071 | 0.7273 | 0.7273 |
| -50 db | 0.5455 | 0.5556 | 0.4545 |

**Table 4- 4 Overall accuracies of classification with DWT
for the two-class systems**

| SNR | Nearest Mean | Maximum Likelihood | Nearest Neighbor |
|--------|--------------|--------------------|------------------|
| 50db | 0.9394 | 0.9394 | 0.9697 |
| 20 db | 0.9293 | 0.9394 | 0.9697 |
| 10 db | 0.9293 | 0.9293 | 0.9697 |
| 0 db | 0.9293 | 0.9293 | 0.9697 |
| -10 db | 0.899 | 0.9091 | 0.9596 |
| -20 db | 0.697 | 0.7475 | 0.7576 |
| -50 db | 0.5354 | 0.5455 | 0.4545 |

**Table 4- 5 Overall accuracies of classification with DWT
for the multi-class systems**

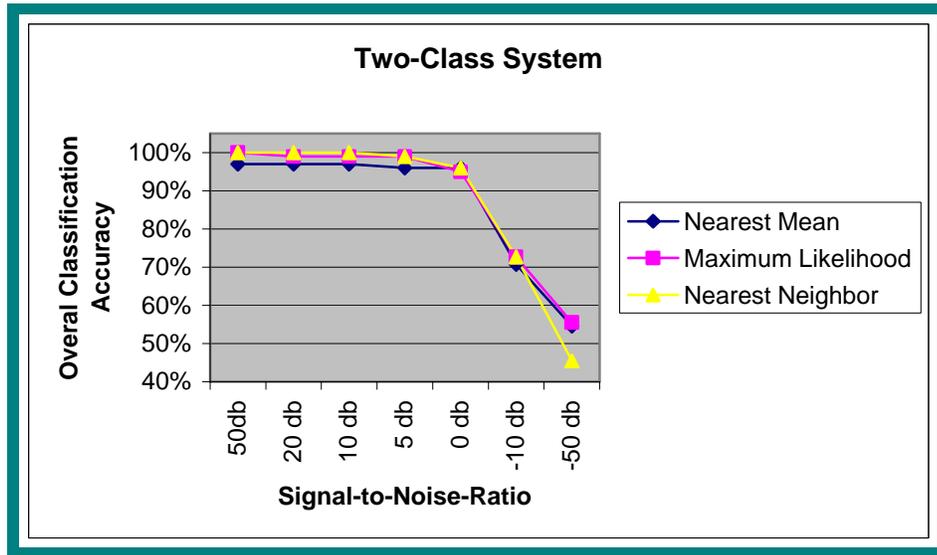


Figure 4- 2 DWT-based feature extraction using Haar mother wavelet for two-class system

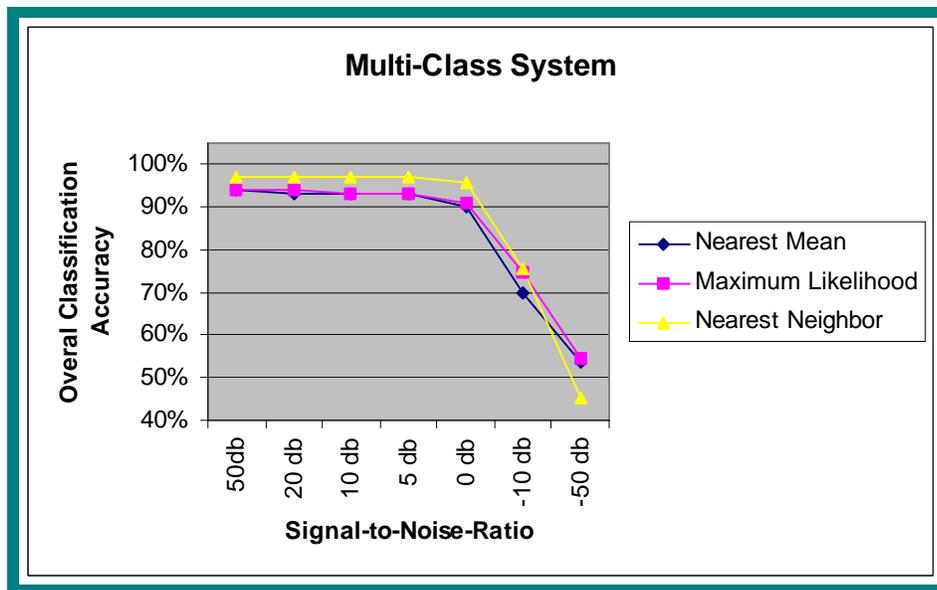


Figure 4- 3 DWT-based feature extraction using Haar mother wavelet for multi-class system

Figures 4-4 and 4-5 show a comparison of the five different feature extraction methods with respect to SNR. The Haar mother wavelet is used with DWT method, and the maximum likelihood classifier is utilized. It is apparent that the wavelet and cepstral approaches outperform the FFT and DCT methods. As long as the SNR is ≥ 0 dB, the DWT and RCT methods detect accidents with an accuracy $\geq 95\%$ and $\geq 90\%$ for the two-class and the multi-class system, respectively. This result is very promising considering the fact that the 0dB case is a very conservative scenario (the crash sound and the background traffic noise are equally loud). The RCT outperforms the DWT method in most cases. However, the DWT method is much more computationally efficient than the RCT. That is, the RCT method provides superior classification accuracies but at a very high computational cost. Also, note that for the two-class system with SNR greater than 0dB, the RCT and DWT methods perform equally well. This is significant since the goal of the system is to detect traffic accidents at intersections, and a SNR greater than 0dB is more realistic. Also, consider the case of implementing the classification algorithms in a real-time system. The DWT method, especially when using the Haar mother wavelet, would be a much more practical choice.

Methods used:

- Feature extractor: DWT(Haar), FFT, DCT, RCT, MCT
- Statistical classifier: Maximum Likelihood Classifier
- Leave-one-out testing
- Data normalize with LDA

| SNR | DWT (Haar) | FFT | DCT | RCT | MCT |
|--------|------------|--------|--------|--------|--------|
| 50db | 1 | 0.9293 | 0.9394 | 1 | 1 |
| 10 db | 0.9899 | 0.899 | 0.9091 | 1 | 0.9899 |
| 5 db | 0.9899 | 0.8687 | 0.899 | 1 | 0.9495 |
| 0 db | 0.9495 | 0.8384 | 0.8182 | 0.9899 | 0.8384 |
| -10 db | 0.7273 | 0.7273 | 0.7071 | 0.8384 | 0.6364 |
| -50 db | 0.5556 | 0.6566 | 0.697 | 0.4949 | 0.4242 |

Table 4- 6 Maximum likelihood classification accuracies for various feature extraction methods for the two-class system

| SNR | DWT (Haar) | FFT | DCT | RCT | MCT |
|--------|------------|--------|--------|--------|--------|
| 50db | 0.9394 | 0.8788 | 0.798 | 0.9899 | 0.9798 |
| 10 db | 0.9394 | 0.8687 | 0.7475 | 0.9899 | 0.9576 |
| 5 db | 0.9293 | 0.8687 | 0.7273 | 0.9899 | 0.899 |
| 0 db | 0.9091 | 0.7576 | 0.6263 | 0.9899 | 0.8081 |
| -10 db | 0.7475 | 0.6465 | 0.5859 | 0.8485 | 0.5758 |
| -50 db | 0.5455 | 0.5859 | 0.404 | 0.4949 | 0.4343 |

Table 4- 7 Maximum likelihood classification accuracies for various feature extraction methods for the multi-class system

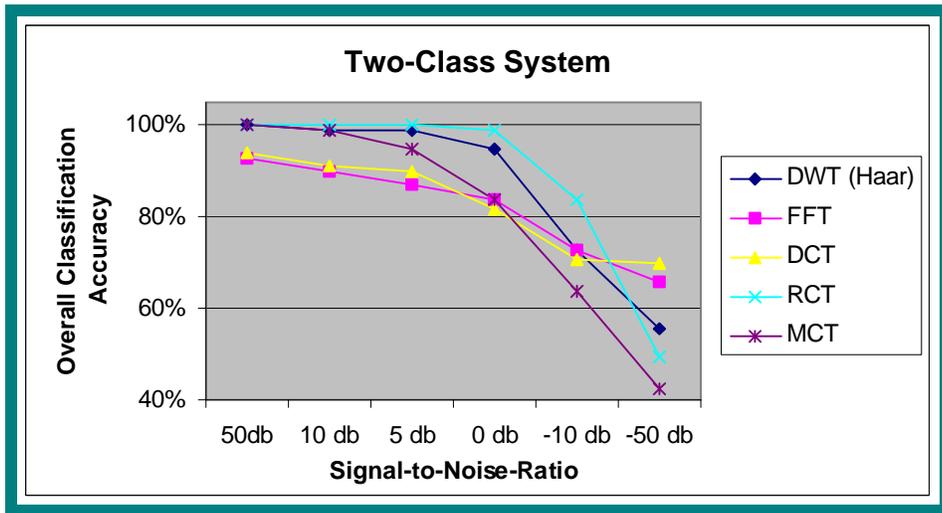


Figure 4- 4 Maximum likelihood classification accuracies for various feature extraction methods for the two-class system

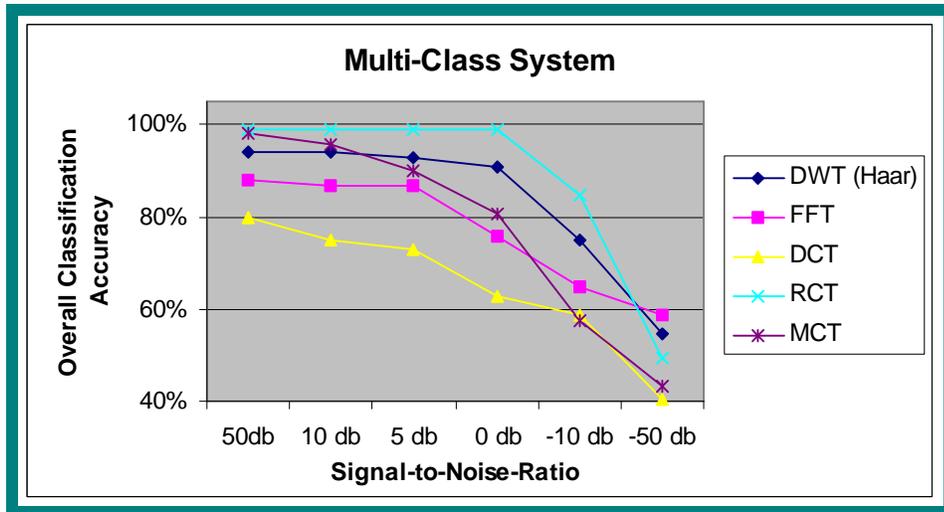


Figure 4- 5 Maximum likelihood classification accuracies for various feature extraction methods for the multi-class system

From the above results analysis, we can find out that the DWT method would be preferred, particularly when using the Haar mother wavelet and the maximum likelihood classifier. The optimal combination was tested using the data collected from the intersections of Jackson MS. Table 4-8 and Table 4-9 show the two-class and multi-class classification accuracies of the DWT (Haar) feature extraction method and the maximum likelihood statistical classifier. In this test, the audio signals were manipulated to model various ambient noise conditions. The SNR's were varied from -50dB to 50dB . Note that the -50dB case represents a scenario that the crash audio signal is very low in amplitude (quiet) as compared with the non-crash audio signals (loud). The 50dB case represents a scenario that the crash audio signal is very high in amplitude (loud) as compared with the non-crash audio signals (quiet). The 0dB case models a scenario that the crash signals and non-crash signals have the same volume. From Tables 4-8 and 4-9, we can see that the classification accuracies are about 95% when SNR is greater than 0dB .

Methods used:

- Feature extractor: DWT(Haar)
- Statistical classifier: Maximum Likelihood Classifier
- Leave-one-out testing
- Data normalize with LDA

| SNR | Crash | Non-crash | Overall Accuracy | Confidence Interval (95%) |
|--------|--------|-----------|------------------|---------------------------|
| 50db | 0.9630 | 0.9444 | 0.9495 | 0.0361 |
| 20 db | 0.9630 | 0.9444 | 0.9495 | 0.0361 |
| 10 db | 0.9630 | 0.9444 | 0.9495 | 0.0361 |
| 0 db | 0.9630 | 0.9167 | 0.9293 | 0.0423 |
| -10 db | 0.7770 | 0.8750 | 0.8485 | 0.0591 |
| -20 db | 0.3304 | 0.7500 | 0.6468 | 0.0788 |
| -50 db | 0.2963 | 0.6528 | 0.5586 | 0.0819 |

Table 4- 8 Classification Results with DWT and maximum likelihood classification for the two-class systems

| SNR | Crash | Non-crash | Pile-drive | Overall Accuracy | Confidence Interval (95%) |
|--------|--------|-----------|------------|------------------|---------------------------|
| 50db | 0.0963 | 1.0000 | 0.6000 | 0.9495 | 0.0361 |
| 20 db | 0.0963 | 1.0000 | 0.6000 | 0.94954 | 0.0361 |
| 10 db | 0.0963 | 1.0000 | 0.6000 | 0.9495 | 0.0361 |
| 0 db | 0.8519 | 1.0000 | 0.7000 | 0.8788 | 0.0538 |
| -10 db | 0.6296 | 0.9576 | 0.8000 | 0.8485 | 0.0591 |
| -20 db | 0.2222 | 0.8387 | 0.6000 | 0.6465 | 0.0788 |
| -50 db | 0.0370 | 0.7581 | 0.6000 | 0.5455 | 0.0821 |

Table 4- 9 Classification Results with DWT and maximum likelihood classification for the multi-class systems

CHAPTER 5.SUMMARY AND FUTURE ORIENTATIONS

5.1 Summary of Testing Results and Recommendations

A system was designed and tested to use audio signals for automated detection of traffic accidents at intersections. Various feature extraction methods were investigated, including techniques based on the DWT, FFT, DCT, RCT, and MCT. As well, various statistical classifiers were investigated, including nearest mean, maximum likelihood, and nearest neighbor. The system was tested on recorded audio signals of normal traffic and traffic accidents, including data collected from Jackson, MS and Starkville, MS. The testing results showed that:

- Among three statistical classifiers, maximum likelihood classifier produced the best results.
- Among five feature extractors, in terms of overall classification accuracy, the RCT feature extraction method worked best. However, when the audio signal's SNR was greater than 0dB, the RCT and DWT methods produced comparable accuracies, $\approx 99\%$. Moreover, the DWT approach was much more computationally efficient than the RCT method, so DWT would be preferred.
- Within the DWT approach, seven types of mother wavelets were tested and compared. The DWT method using the Haar mother wavelet would be a much more practical choice.

In summary, the wavelet-based features extractor in combination with the maximum likelihood classifier is the optimum design. The system is computationally inexpensive relative to the other methods investigated, and the system consistently results in accident

detection accuracies ranging from 95% to 100%, when the audio signal has a signal-to-noise-ratio of at least 0 decibels. The results showed that the method is capable of effectively performing crash and non-crash classification of acoustic signals.

High accident detection accuracy has been achieved through this method, but the process has not been carried out in real time. Data was collected at intersections and then processed and analyzed later in the lab. In order for the automated accident detection system to be a meaningful component of the Intelligent Transportation Systems (ITS), the algorithms must be implemented such that real-time detection can be completed. “Real-time Accident Detection System” should be developed.

5.2 Future Work

The next phase of this project will be focused on real-time testing of the algorithms that were developed and selected during the current phase of algorithm development. Ideally, real-world crash signal should be obtained and testing of the system be conducted in real-time. Issues such as computation time and accuracy will be evaluated, and potential new algorithms will be also evaluated for the purpose of real-time application.

Another very important task for the next phase is the determination of system architecture for the real-time system. The intersection accident detection system should have real-time capabilities in data recording, signal processing, and data communication. The real-time accident detection system can be implemented using one of the following two system architectures:

(1) Decentralized architecture

(2) Centralized architecture.

The main difference between the two approaches is where the received acoustic signals will be processed and what data will be transmitted to the traffic management center.

The decentralized architecture has particularly high operational requirements for the active sensors. Almost all of the signal processing is completed at the intersection. However, the volume of the data need to be transmitted from the acoustic sensor to the central server is very low. Therefore, this approach has a very low requirement for the performance of the communication network and the central server. On the contrary, the centralized design has particularly high performance requirements for the communication network and the central server. A large volume of unprocessed signals need to be transmitted from the acoustic sensor through the communication network, and then be processed simultaneously by the central server. The pros and cons of the two designs will be compared based on cost, performance, reliability, and scalability.

Transmission of data between intersections and the traffic management center is an important component of the Real-time Intersection Accident Detection System. Data transmission could be done using telephone lines via modems, using existing owned cabling, or using wireless communication systems. The choice will depend on the design structure and capacity requirements, as well as the ability to use existing communication networks. Cost, performance, reliability, and scalability are again the criteria upon which the selection will be made.

The final recommendations on system architecture and related implementation issues, including communication configurations, for the real-time system will be investigated during the next phase of the project. The real-time testing will be carried out in order to evaluate the feasibility of a real-time system.

REFERENCES:

- [1] TTI 1999 Urban Mobility Report, Texas Transportation Institute. 1999.
- [2] TTI 2002 Urban Mobility Report-Summary of the Austin Data, Texas Transportation Institute, June 20, 2002
- [3] INCIDENT MANAGEMENT PROGRAM BACKGROUND, Spring 2002.
<http://kdot1.ksdot.org/public/kdot/kcmetro/pdf/Ch1.pdf>. Accessed Dec. 20, 2002
- [4] J. M. McDermott, "Incident Surveillance and Control on Chicago-Area Freeways," Special Report 153: Better Use of Existing Transportation Facilities. TRB, National Research Council, Washington, D.C., 1975,pp. 123–140.
- [5] H.J. Payne, E.D. Helfenbein, and H.C. Knobel, Development and Testing of Incident Detection Algorithms: Volume 2-Research Methodology and Detailed Results, Report No. FHWA-RD-76-20, Washington, D.C., Federal Highway Administration, 1976.
- [6] David A. Whitney and Joseph J. Pisano, TASC, Inc., Reading, Massachusetts. "AutoAlert: Automated Acoustic Detection of Incidents" December 26, 1995
- [7] Marc Solomon, NSF Research Fellow, A Review of Automatic Incident Detection Techniques, Transportation Center, Northwestern University. August 16, 1991
- [8] Dr. Peter T. Martin, Joseph Perrin, Blake Hansen, Ryan Kump, Dan Moore, INCIDENT DETECTION ALGORITHM EVALUATION, Prepared for Utah Department of Transportation, University of Utah, March 2001
- [9] A Review of Current and Future Data Requirements and Detector Technologies and the Implications for UTMC" June 2000.
- [10] S. Burrus, R. Gopinath, H. Guo, Introduction to Wavelets and Wavelet Transforms: A Primer, 1/e, Prentice-Hall, New Jersey, 1998.
- [11] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, PA. 1992.
- [12] G. Strang and T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, 1996.

-
- [13] Sophocles J. Orfanidis, Introduction To Signal Processing, Rutgers University, Prentice Hall, Upper Saddle River, New Jersey 07458, 1996.
- [14] Andrew Bruce, Hong-ye Gao, Applied Wavelet Analysis with S-PLUS. Mathsoftm Inc. 1996
- [15] X. Huang, A. Acero, H. Hon, Spoken Language Processing: A Guide to Theory, Algorithm, and System Development, pp. 306-318, Prentice-Hall, 2001.
- [16] R.Duda, P.Hart, D.Stork, Pattern Classification, Wiley-Interscience, 2001.
- [17] A.V. Oppenheim, R.W. Schaffer, Discrete-Time Signal Processing, 2nd ed., pp. 589-598, Prentice-Hall, 1999
- [18] C. Harlow and Y. Wang, "Automated Accident Detection," Proc. Transportation Research Board 80th Annual Meeting, pp 90-93, 2001.
- [19] L. D. Fisher and G. V. Belle, Biostatistics, A Methodology for the Health Sciences, John Wiley & Sons, Inc., 1993.